Format of the data portal for the C20C Detection and Attribution project

[1]Dáithí Stone (dstone@lbl.gov)

This document provides details on the categories and labels for use in the earthsystemgrid data portal of the C20C Detection and Attribution project. This is largely consistent with the CF (NetCDF Climate and Forecast Metadata Convention) standard.

Location

The data portal is located at:
`http://esg.nersc.gov`

Classification

Project: *C20C Detection and Attribution*

Institute: The institution performing the experiments, preferably a short form.
Example: *LBNL* (for Lawrence Berkeley National Laboratory)

Model: The climate model used for the simulations, including an indication of version if different versions of a model are run or if a model is run a different resolutions.
Example: *CAM5.1-2degree* (for CAM5.1 run at $\sim$2-degree horizontal resolution)

Experiment family: The label for the scenario being used.
For the core experiment this will either be *All-Hist* for the "real-world" simulations or "Nat-Hist" for the simulations of the world that might have been had humans never interfered with the climate.
See Table 1.

Experiment: The label for the estimate of the scenario being used, or equivalently the sub-scenario. For instance, for the "Nat-Hist" scenario one estimate will use the attributable ocean warming estimate from one coupled climate model, while another estimate will use the warming estimate from another coupled model. This label will also be used to distinguish between ensembles of simulations driven with different original forcing datasets, for instance different estimates of the solar forcing.
See Table 2.

Time Frequency: *Fixed*, *Monthly*, *Daily*, or *3-Hourly*

Realm: The part of the climate system of interest.
Either *atmosphere* or *land*. Additional values may be generated to classify derivatives calculated from model output in subsequent analyses, for instance extreme value indices or agricultural indices.

Variable: The CF label of the variable.
Example: *tas* (for near-surface air temperature).
See document "Requested metadata for output from the C20C Detection and Attribution project".

Ensemble: An identifier of a simulation within the Experiment's ensemble of simulations.
Example: *run007*.

Label	**Description**
All-Hist	Historical (1960-2011) including changes in "all" known external forcings (anthropogenic and natural)
Nat-Hist	Historical (1960-2011) including changes in natural external forcings only

Table 1: Labels of options for the **Experiment_family** classifier for the core experiment of the C20C Detection and Attribution project.

Experiment_family label	**Experiment label**	**Description**
All-Hist	v1	Using a certain set of external forcings (e.g. HadISST2 SST and SIC, and Lean (2005) solar forcing)
	v2	Like "v1" but using a different set of external forcings (e.g. NOAA OI.v2 SST and SIC, and Hoyt and Schatten (1993) solar forcing)
Nat-Hist	CanESM2-p25-v1	Using a certain set of original external forcings (e.g. HadISST2 SST and SIC) and altering SSTs and SICs according to a certain estimate of the attributable warming due to anthropogenic activites (in this case for the 25th percentile of the estimate from optimal fingerprinting using CMIP5 simulations of the CanEMS2 AOGCM)
	HadGEM2-ES-p75-v1	Like "CanESM2-p25-v1" but altering SSTs and SICs according to a different estimate of the attributable warming due to anthropogenic activites (in this case for the 75th percentile of the estimate from optimal fingerprinting using CMIP5 simulations of the HadGEM2 AOGCM)
	CanESM2-p25-v2	Like "CanESM2-p25-v1" but using a different set of original external forcings (e.g. NOAA OI.v2 for SST and SIC)

Table 2: Labels of possible entries for the **Experiment** classifier for the core experiment of the C20C Detection and Attribution project. The component of the "Nat-Hist" label designating the attributable ocean warming estimate used (e.g. "CanESM2-p25") will be provided with the SST and SIC dataset for running the simulations; the version component of the label (e.g. "v2") will need to be determined locally by the modelling institution.

File names

It would be useful if filenames took the following form, essentially following the format used by CMIP5 but with the <experiment_family> field added:

`<variable>_<frequency>_<model_id>_<experiment_family>_<experiment>_run<run_id>_<start>-<end>.nc`

<start> and <end> should be numerical and of the form YYYYMM, for instance “196001” for January 1960.

For example, for near-surface air temperature (tas), averaged monthly (mon), from model CAM5.1-2degree, in simulation #15 run under scenario All-Hist, with data from January 1960 (196001) through December 2011 (201112), the filename would be:

`tas_mon_CAM5-1-2degree_All-Hist_v1_run15_196001-201112.nc`

Note that CAM5.1-degree has been converted to “CAM5-1-2degree” for the filename: it is convenient to leave dots out of the filename except in the “.nc” at the end.